1 **Title:** Functional Redundancy in Ocean Microbiomes Controls Trait Stability

2

3 **Authors:** Taylor M. Royalty[1] and Andrew D. Steen[1,2]

4 **Affiliations:** University of Tennessee Departments of Microbiology[1] and Earth and Planetary

5 Sciences[2]

Abstract:

Advances in nucleic acid sequencing technology have revealed that, in many microbial ecosystems, the same ecosystem function or trait is performed by multiple species or taxa. Theory, developed in the context of macroecology, predicts that communities with high functional redundancy are less likely to lose functions due to species extinction compared to communities with low functional redundancy. It is not clear whether this is the case for microbial communities, particularly on the landscape scale. In part, the lack of quantitative measures for functional redundancy in microbial ecosystems has been prohibitive in addressing this question. We recently proposed a quantitative functional redundancy metric, contribution evenness, which measures how evenly taxa in a community contribute to an ecosystem function or trait. Using transcriptomes deposited in the Ocean Microbial Reference Gene Catalog (OM-RGC.v2), a catalog of genes and transcripts sequenced by the *TARA Ocean* expedition, we quantified the functional redundancy for 4,314 KEGG Orthologs (KOs) across 124 marine sites. Functional redundancy was highly correlated with a latent variable reflecting few ocean physiochemical parameters and was systematically higher at the poles than in non-polar regions. Functional richness β-diversity among non-polar sites was higher than that among polar sites, indicating that microbial ecosystem functions are more similar among polar sites than among non-polar sites. These observations combined provide evidence that functional redundancy influences microbial ecosystem function stability on spatiotemporal scales consistent with surface ocean mixing. We suggest that future changes in ocean physiochemistry will likely influence this stability.

Main Text:

The ability to resolve complete genomes in microbial communities via contig binning and single-cell genome sequencing has revealed that metabolic functions, or traits, are often present in multiple taxa within a microbiome (*1–4*). The practical consequences of functional redundancy, in terms of community composition and function, are not well understood for microbial communities. Theoretical predictions, developed in the context of macroecosystems, suggest that functional redundancy can buffer a community against function loss due to species extinction (*5–9*). Foundational work on grassland ecosystems validated these theoretical predictions by showing that plant community biomass stability increased with higher functional redundancy (*10*). Mesocosm experiments provide empirical evidence that function replication among species may also buffer microbial ecosystem function from species loss. For instance, microbial density and biomass was less variable for assemblages with redundant populations occupying different trophic guilds in a food web (*11*), and reduction-oxidation conditions in sediment communities were more stable with increased bacterial diversity and niche overlap (*12*). Although existing theoretical and empirical evidence suggests functional redundancy may enhance ecosystem function stability, small-scale and mesocosm experiments with microbial communities may not reflect the most important processes on ecosystem scales (*13*). Establishing whether functional redundancy stabilizes traits is important for understanding how environmental fluctuations, most notably climate change, may influence microbially-mediated biogeochemistry stability.

We used contribution evenness (*14*) to measure functional redundancy for KEGG ortholog (KO)-annotated genes from 180 and 124 *TARA Oceans* metagenomes and metatranscriptomes, respectively (*15*). Contribution evenness uses genes (contribution evenness

with respect to genes) or transcripts (contribution evenness with respect to transcripts) as traits (*16*, *17*) and measures the evenness how evenly taxa in community contribute to these traits as a whole. The *TARA Oceans* dataset was appealing because sample collection was largely standardized for biological sequences and physiochemistry while simultaneously covering a diverse range of sample sites. After calculating contribution evenness for metagenomes (Fig. 1A) and metatranscriptomes (Fig. 1B), we found that functional redundancy varied substantially among metabolism type and sites. In total, individual gene and transcript functional redundancy spanned over three orders of magnitudes and exhibited differences in distributions among the selected metabolism types. Furthermore, distribution of transcript abundance within ocean microbiomes correlated well with the distribution of gene abundance at the same site (Fig. 1C; R^2=0.45, p<0.001). That is, risk of transcription loss is strongly influenced by the fate of genes. This relationship was not surprising considering previous reports that metagenome gene abundances strongly influenced metatranscriptome transcript abundances across *TARA Oceans* microbiomes (*15*). Although the nature of these observations is similar, the ecological implication is different. Specifically, the observation of Salazar et al. shows correlation in absolute genes abundance and transcription abundance, whereas our observation indicates that when genomic traits that are more evenly distributed across a community, transcription is also more evenly distributed across a community.

In order to better understand the relationship between environmental characteristics and functional redundancy, we used redundancy analysis (RDA) to model the functional redundancy of each KO at each site as a function of the seven environmental parameters: salinity, depth, and concentrations of nitrate, phosphate, and silicate ions, oxygen, and chlorophyll A (Fig. 1). The first canonical axis of this model explained 19.8% of the total functional redundancy variance in

4,314 transcripts annotated as KOs. An ANOVA demonstrated that the first canonical axis explained significantly more variability in KO functional redundancy than a null model (p<0.01) (*18*). Although the total explanatory power for individual KOs was low, we found that the average functional redundancy of all KOs at each TARA Oceans site was accurately predicted very by the first canonical axis alone (93.2% of total variance explained; ordinary least squares regression.

Factor loadings of this first canonical axis were dominated positively by oxygen and chlorophyll A and negatively by sample depth and salinity (Fig. 2B). Thus, this canonical axis appears to be a proxy for photosynthesis rate, or the extent to which an environment is dominated by copiotrophs versus oligotrophs. The strong predictive power of the photosynthesis was not surprising given that carbon export mediated by primary productivity is known to be a good predictor for community composition and genomic composition in the ocean (*19*). Higher functional redundancy was positively correlated with higher photosynthesis (i.e., positive scores). We compared the performance of our factor-based model to a best subset (based on an AIC criteria) OLS regression using the original seven physiochemical variables (NO_3^-, PO_4^{3-}, salinity, depth, O_2, Si, and ChlA) that generated our factor (Fig. 2C). In contrast to the factor model, the best subset model (O_2, Si, and depth) explained only 32.4% of the variance in mean functional redundancy.

The high explanatory power by the environmental data justified extrapolating our factor model onto a global scale. Utilizing global predictions of ocean nutrient concentrations, we calculated factor scores at a 0.25°x0.25° resolution and predicted functional redundancy across Earth's oceans using the derived coefficient from our OLS regression of mean functional redundancy versus the first canonical axis of the RDA. Our model predicts that functional

redundancy is highest in polar regions and near river outflows, and lowest in subtropical gyres (Fig. 3A). Variance was highest in polar regions, coastal regions, and river outflows (Fig. 3B). The transition from high to low functional redundancy between polar and non-polar latitudes was consistent with previously reported ecological boundaries for ocean microbiomes, where the transition from non-polar to polar latitudes corresponded to compositional changes in metatranscriptomes and metagenomes (*15*).

We hypothesized that this latitudinal gradient in functional redundancy would correlate with trait stability, or intra-regional (i.e., polar and non-polar regions) variance in the absence and/or presence of traits. Again, theory predicts functional redundancy can buffer communities from changes in functional composition when multiple taxa perform said functions (*5–9*, *14*). Regions with higher functional redundancy should have less inter-site variance in trait composition functional redundancy promotes stability. One approach to test this idea is to compare functional redundancy to functional richness β-diversity, where regional measures of functional richness β-diversity should inversely relate to functional redundancy. We performed a bootstrap subsampling analysis on polar and non-polar metatranscriptomes and measured the functional richness β-diversity (*20*) within the respective groups. Functional richness β-diversity was significantly higher at lower latitudes based on an ANOVA ($p<0.001$) (Fig. 4). We take this as evidence that functional redundancy supports higher trait stability in ocean microbiomes.

Trends suggest that major ocean currents are slowing (*21*), oxygen minimum zones are expanding (*22*), and there will be future changes in regional primary productivity (*23*). Notably, ocean currents mediate microbial dispersion across Earth's oceans. Reduction in dispersion rates will theoretically lower and increase α- and β-diversity in metacommunities, respectively (*24*), which should result in higher heterogeneity in microbially-mediated biogeochemistry. With

respect to oxygen and primary productivity, these variables are important in our factor model and suggests future changes in these parameters will influence ocean microbiome functional redundancy. Although we demonstrated that surface ocean microbiome functional redundancy and trait stability relate to one another, new questions arose about how environment directly influences this functional redundancy, and thus, trait stability. This will be a point of future exploration.

Materials and Methods:

TARA Oceans Genes and Environmental Data

We downloaded the entire Ocean Microbial Reference Gene Catalog v2 (OM-GRC.v2) on 1 Feb 2021 from EMBL-EBI (BioStudy: S-BSST297) and the environmental data from https://doi.org/10.5281/zenodo.3473199 (*15*). With OM-RGC.v2, we evaluated gene and transcript profiles isolated from biological sequences collected on filters with size ranges. Each profile corresponded to filtration collected from 100L of seawater in the size range 0.22-1.6 μm or 0.22-3.0 μm (*26*). Profiles were filtered for sequences annotated to KEGG orthologs (KO)(*27*). Our preference of KO annotations versus cluster of orthologous groups (COGs) or gene clusters (GC) pertains to KOs mapping to specific metabolic processes. We interpreted length-normalized short-read mapping frequencies as abundances. These values are synonymous to the quantity, or abundance, of any sequence in a metagenome or transcriptome, given a fixed sequencing effort. Gene abundances for KO single copy marker genes (list below) annotated to the same genus were averaged to estimate the abundance of a given genus. Our selection of genus level analysis was to mitigate noise associated with analyzing functional redundancy of ASVs, which is theoretically possible. To allow for meaningful comparisons of functional redundancy at different sites, we rarified sequencing effort at each site. Relative abundances

were then used as weights during random sampling, with replacement. Each site was sampled 3358334 and 472163 times for gene and transcript profiles, respectively. These values corresponded to the lowest sampling effort among all metagenomes and transcriptomes, respectively.

KO Functional Redundancy and Diversity

Functional redundancy was calculated for each KO, at each site. Functional redundancy was calculated as contribution evenness (*14*), a measure of how evenly taxa contribute to an ecosystem level process (e.g., photosynthesis, sulfate-reduction, glycolysis, etc.). Here, we treat KOs as proxies for metabolic potential. To determine how evenly taxa contribute to an ecosystem-level process, taxa abundances are multiplied onto taxa-specific trait levels (genes per cell and transcripts per cell) and normalized to sum to one. This distribution reflects the relative contribution to a metabolism process. As the OM-RGC.v2 gene profiles reflect abundances in metagenomes and transcriptomes, we did not need to multiple taxa abundances onto taxa-specific trait levels. To obtain KO functional redundancy, the effective number of "contributors" is calculated from profiles (after rarifying) using effective diversity (*28*, *29*). The effective number of contributors is then normalized by taxa richness and adjusted so that the minimum and maximum evenness equal 0 and 1, respectively (*30*). As such, contribution evenness is calculated as:

$$R = \frac{D_q - 1}{S - 1} \quad (1$$

where R is functional redundancy, D_q is the effective number of contributors (of the q^{th} diversity order), and S is taxa richness. Consistent with Royalty & Steen (2021), we used a diversity order q=0.5 as this diversity order corresponded to minimizing the mean absolute error between

functional redundancy and trait resilience to taxa extinction. Here, taxa richness was calculated using KO single copy marker genes. The richness of taxonomic families was determined by determining number of the average abundance of K06942, K01889, K01887, K01875, K01883, K01869, K01873, K01409, K03106, and K03110 was greater than zero Salazar et al. (2019). Last, each KO functional redundancy, from all sites, was normalized by the KO's respective variance (standard deviation) in functional redundancy. This variance was calculated using the KO's functional redundancy from across all sites. Normalizing the functional redundancy of individual KOs by their variance was necessary as mean functional redundancy is correlated to KO functional redundancy variance (Figure S1). Low functional redundancy does not necessarily imply that a function is ecologically unimportant. This is particularly true for traits with low functional redundancy that strongly correlate to phylogeny. Such examples include *Thaumarchaeota* and diazotrophs which act as the primary nitrifiers and nitrogen fixers in the ocean, respectively (*14*). To the contrary, small changes in functional redundancy for traits with nominally low functional redundancy might reflect key ecological processes, and thus, comparisons of community functional structure should account for the differences in functional redundancy magnitude. Thus, changes in KO functional redundancy reflected changes with respect to the KO's observed variance while not reflecting the KO's order of magnitude, which spanned five orders of magnitude.

Diversity was calculated as effective richness (*28*, *29*). The abundance of each genus was determined by taking the average abundance of KO single copy marker genes (see above). The average abundance of KO single copy marker genes were normalized by the total average KO single copy marker gene abundance. Effective richness was then calculated using a diversity order q=1, which corresponds to taking the exponential of Shannon's Index. This value

corresponds to the minimum number of evenly abundant families necessary to obtain an observed entropy.

Global Modeling of Functional Redundancy

We sought to characterize spatial trends in the functional structure of microbial communities. To do this, we performed redundancy analysis using the rda function from the R package, vegan (*31*). In this analysis, the functional structure of microbial communities was treated as a response variable and physio-chemical variables were treated as predictor variables. For sites where a given KO was absent, the KO was assigned a value of zero. This situation occurred ~13.6% and 18.5% of the time for gene and transcript profiles, respectively. For this analysis, we used salinity (PSU), nitrate (mmol m^{-3}), phosphate (mmol m^{-3}), oxygen (mmol m^{-3}), chlorophyll A (mg m^{-3}), depth (m), and silicate (mmol m^{-3}). We chose these variables as they were *in situ* measurements (*26*) and could scale with models that predict global ocean chemistry. We imputed missing data, as removing incomplete cases can bias datasets (*32*). Random forest imputation was performed using the missForest function, from the R package, missForest (*33*). Next, each variable was converted into a normal distribution using a boxcox transformation via the boxcox function in the R package, MASS (*34*). Variables were then centered to have a mean of zero. The significance of each variable as well as the first two canonical axes was verified using the anova.cca function in the R Package, vegan (*31*).

The first canonical axis derived from the redundancy analysis was used to predict mean transcript functional redundancy from the individual metatranscriptomes (sites) using OLS regression. This model was compared to an OLS regression using salinity, nitrate, phosphate, oxygen, chlorophyll A, depth, and silicate as predictors of mean transcript functional redundancy. Similar to before, missing data were imputed using missForest (*33*), variables were

transformed with a boxcox transformation (*34*), and variables centered so the mean distribution was zero prior to regression. The best predictor subset was determined using the regsubsets function from the R package, leaps (*35*). The criteria defining the best model was minimizing AIC.

The monthly-averaged data products spanning from Jan 2013 to Dec 2018 for GLOBAL_REANALYSIS_BIO_001_029 and GLOBAL_REANALYSIS_PHY_001_031 were downloaded from https://marine.copernicus.eu/ on 7 May 2021. Data products had a grid resolution of 0.25°x0.25°. The OLS regression using the first canonical axis as a predictor substantially outperformed the best subset OLS when predicting mean transcript functional redundancy. As such, we converted the predicted data product chemistry for each grid cell into a score for the first canonical axis. Then, the mean transcript functional redundancy was predicted for each grid cell, for each month, utilizing the coefficient derived from the canonical axis OLS regression. The median, 5^{th} percentile, and 95^{th} percentile of mean functional redundancy was taken across the six-year window. Variance was measured as the difference between the 95^{th} and 5^{th} percentile divided by the median.

Comparing Metatranscriptome α-Diversity, β-diversity, and γ-Diversity Between Polar and Non-polar Latitudes

We performed a bootstrap analysis to compare the α-diversity, β-diversity and γ-diversity among polar and non-polar sites. In brief, we randomly sampled, with replacement, 10 epipelagic (<200m deep) polar (n=38) and non-polar (n=74) metatransciptomes, respectively. The α-diversity and γ-diversity were calculated as the mean richness and total richness, respectively, among the 10 randomly sampled sites. Then, we solved for β-diversity (*20*) for both the polar and non-polar metatranscriptomes such that:

$$D(H_\beta) = \frac{D(H_\gamma)}{D(H_\alpha)} \quad (2$$

We sampled in this fashion for 1,000 iterations prior to comparing α-diversity, β-diversity and γ-diversity.

Acknowledgements:

Funding for this project was provided by the Department of Energy, Office of Science, Office of Biological and Environmental Research (DE-SC0020369) to A.D.S.

Competing Interests:

We have no competing interests.

Data and Materials Availability:

The Ocean Microbial Reference Gene Catalog v2 (OM-GRC.v2) is accessible from EMBL-EBI (BioStudy: S-BSST297). *TARA Oceans* site metadata was downloaded from https://doi.org/10.5281/zenodo.3473199. The hindcast data products, GLOBAL_REANALYSIS_BIO_001_029 and GLOBAL_REANALYSIS_PHY_001_031, are accessible from https://marine.copernicus.eu/. All code and processed data is available at https://github.com/taylorroyalty/tara_ocean_fr.

References and Notes:

1. S. Louca, M. F. Polz, F. Mazel, M. B. N. Albright, J. A. Huber, M. I. O'Connor, M. Ackermann, A. S. Hahn, D. S. Srivastava, S. A. Crowe, M. Doebeli, L. W. Parfrey, Function and functional redundancy in microbial systems. *Nat. Ecol. Evol.* **2**, 936–943 (2018).

2. T. M. Royalty, A. D. Steen, Quantitatively Partitioning Microbial Genomic Traits among Taxonomic Ranks across the Microbial Tree of Life. *mSphere*. **4**, e00446-19 (2019).

3. J. B. H. Martiny, S. E. Jones, J. T. Lennon, A. C. Martiny, Microbiomes in light of traits: A phylogenetic perspective. *Science (80-.).* **350**, aac9323 (2015).

4. T. O. Delmont, C. Quince, A. Shaiber, Ö. C. Esen, S. T. Lee, M. S. Rappé, S. L. MacLellan, S. Lücker, A. M. Eren, Nitrogen-fixing populations of Planctomycetes and Proteobacteria are abundant in surface ocean metagenomes. *Nat. Microbiol.* **3**, 804–813 (2018).

5. S. Naeem, Species Redundancy and Ecosystem Reliability. **12**, 39–45 (1998).

6. J. S. Rosenfeld, M. Mall, Functional redundancy in ecology and conservation. *OIKOS*. **1** (2002).

7. C. R. Fonseca, G. Ganade, Species Functional Redundancy, Random Extinctions and the Stability of Ecosystems. *J. Ecol.* **89**, 118–125 (2001).

8. B. H. Walker, Biodiversity and Ecological Redundancy. *Conserv. Biol.* **6**, 18–23 (1992).

9. K. L. Cottingham, B. L. Brown, J. T. Lennon, Biodiversity may regulate the temporal variability of ecological systems. *Ecol. Lett.* **4**, 72–85 (2001).

10. D. Tilman, P. B. Reich, J. M. H. Knops, Biodiversity and ecosystem stability in a decade-long grassland experiment. *Nature*. **441**, 629–632 (2006).

11. S. Naeem, S. Li, Biodiversity enhances ecosystem reliability. **390**, 507–509 (1997).

12. E. R. Hunting, M. G. Vijver, H. G. van der Geest, C. Mulder, M. H. S. Kraak, A. M. Breure, W. Admiraal, Resource niche overlap promotes stability of bacterial community metabolism in experimental microcosms. *Front. Microbiol.* **6**, 1–7 (2015).

13. C. E. ZOBELL, D. Q. ANDERSON, Observations on the Multiplication of Bacteria in Different Volumes of Stored Sea Water and the Influence of Oxygen Tension and Solid Surfaces. *Biol. Bull.* **71**, 324–342 (1936).

14. T. M. Royalty, A. D. Steen, Contribution Evenness: A functional redundancy metric for microbially-mediated biogeochemical rates and processes. *bioRxiv*, 6 (2021).

15. G. Salazar, L. Paoli, A. Alberti, J. Huerta-Cepas, H. J. Ruscheweyh, M. Cuenca, C. M. Field, L. P. Coelho, C. Cruaud, S. Engelen, A. C. Gregory, K. Labadie, C. Marec, E. Pelletier, M. Royo-Llonch, S. Roux, P. Sánchez, H. Uehara, A. A. Zayed, G. Zeller, M. Carmichael, C. Dimier, J. Ferland, S. Kandels, M. Picheral, S. Pisarev, J. Poulain, S. G. Acinas, M. Babin, P. Bork, E. Boss, C. Bowler, G. Cochrane, C. de Vargas, M. Follows, G. Gorsky, N. Grimsley, L. Guidi, P. Hingamp, D. Iudicone, O. Jaillon, S. Kandels-Lewis, L. Karp-Boss, E. Karsenti, F. Not, H. Ogata, S. Pesant, N. Poulton, J. Raes, C. Sardet, S. Speich, L. Stemmann, M. B. Sullivan, S. Sunagawa, P. Wincker, Gene Expression Changes and Community Turnover Differentially Shape the Global Ocean Metatranscriptome. *Cell*. **179**, 1068-1083.e21 (2019).

16. C. Violle, M.-L. Navas, D. Vile, E. Kazakou, C. Fortunel, I. Hummel, E. Garnier, Let the

concept of trait be functional! *Oikos*. **116**, 882–892 (2007).

17. J. L. Green, B. J. M. Bohannan, R. J. Whitaker, Microbial biogeography: From taxonomy to traits. *Science (80-.).* **320**, 1039–1043 (2008).

18. P. Legendre, J. Oksanen, C. J. F. ter Braak, Testing the significance of canonical axes in redundancy analysis. *Methods Ecol. Evol.* **2**, 269–277 (2011).

19. L. Guidi, S. Chaffron, L. Bittner, D. Eveillard, A. Larhlimi, S. Roux, Y. Darzi, S. Audic, L. Berline, J. R. Brum, L. P. Coelho, J. C. I. Espinoza, S. Malviya, S. Sunagawa, C. Dimier, S. Kandels-Lewis, M. Picheral, J. Poulain, S. Searson, L. Stemmann, F. Not, P. Hingamp, S. Speich, M. Follows, L. Karp-Boss, E. Boss, H. Ogata, S. Pesant, J. Weissenbach, P. Wincker, S. G. Acinas, P. Bork, C. De Vargas, D. Iudicone, M. B. Sullivan, J. Raes, E. Karsenti, C. Bowler, G. Gorsky, Plankton networks driving carbon export in the oligotrophic ocean. *Nature*. **532**, 465–470 (2016).

20. L. Jost, Partitioning Diversity into Independent Alpha and Beta Components. *Ecology*. **88**, 2427–2439 (2007).

21. L. Caesar, G. D. McCarthy, D. J. R. Thornalley, N. Cahill, S. Rahmstorf, Current Atlantic Meridional Overturning Circulation weakest in last millennium. *Nat. Geosci.* **14**, 118–120 (2021).

22. T. Ito, S. Minobe, M. C. Long, C. Deutsch, Upper ocean O2 trends: 1958–2015. *Geophys. Res. Lett.* **44**, 4214–4223 (2017).

23. W. Fu, J. T. Randerson, J. Keith Moore, Climate change impacts on net primary production (NPP) and export production (EP) regulated by increasing stratification and

312 phytoplankton community structure in the CMIP5 models. *Biogeosciences*. **13**, 5151–5170
313 (2016).

314 24. N. I. Wisnoski, M. A. Leibold, J. T. Lennon, Dormancy in metacommunities. *Am. Nat.*
315 **194**, 135–151 (2019).

316 25. P. Legendre, L. Legendre, *Numerical Ecology* (Elsevier, ed. 3, 2012).

317 26. S. Pesant, F. Not, M. Picheral, S. Kandels-Lewis, N. Le Bescot, G. Gorsky, D. Iudicone,
318 E. Karsenti, S. Speich, R. Trouble, C. Dimier, S. Searson, Open science resources for the
319 discovery and analysis of Tara Oceans data. *Sci. Data*. **2**, 1–16 (2015).

320 27. M. Kanehisa, M. Furumichi, M. Tanabe, Y. Sato, K. Morishima, KEGG: New
321 perspectives on genomes, pathways, diseases and drugs. *Nucleic Acids Res.* **45**, D353–
322 D361 (2017).

323 28. L. Jost, Entropy and diversity. *Oikos*. **2** (2006), doi:https://doi.org/10.1111/j.2006.0030-
324 1299.14714.x.

325 29. M. O. Hill, Diversity and Evenness : A Unifying Notation and Its Consequences. *Ecology*.
326 **54**, 427–432 (1973).

327 30. A. Chao, C. Ricotta, Quantifying evenness and linking it to diversity , beta diversity , and
328 similarity. **100**, 1–15 (2019).

329 31. D. Philip, Computer program review VEGAN , a package of R functions for community
330 ecology. *J. Veg. Sci.* **14**, 927–930 (2003).

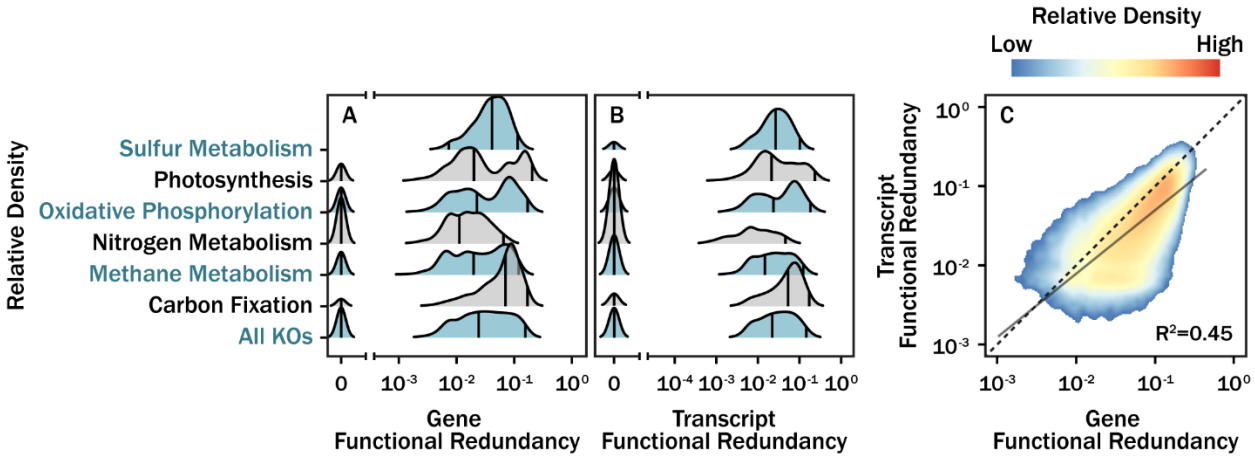
331 32. R. McElreath, *Statistical Rethinking: A Bayesian Course with Examples in R and Stan*
332 (CRC Press, ed. 2nd, 2020).

333 33. D. J. Stekhoven, P. Bühlmann, Missforest-Non-parametric missing value imputation for
334 mixed-type data. *Bioinformatics*. **28**, 112–118 (2012).

335 34. W. N. Venables, B. D. Ripley, Modern Applied Statistics with S (2002).

336 35. T. Lumley, The leaps Package (2004).

337 **Figures:**



338

339 **Fig. 1:** Variation in functional redundancy for KOs across 180 *TARA Oceans* metagenomes (**A**),
340 129 *TARA Oceans* transcriptomes (**B**), and a pairwise comparison of functional redundancy
341 (n≈415,000) for KOs annotated the same at the same site (**C**). Density curves (**A,B**) were
342 generated with a kernel density estimate (gaussian kernel with a bandwidth 0.0803). The vertical
343 black lines correspond to 5th, 50th, and 95th quantiles. The metagenomes and metatranscriptomes
344 functional redundancy regression (**C**) followed a $\log_{10}$-$\log_{10}$ model (solid black line). Prior to
345 regression, all values were offset by 10^{-4} and high leverage data was removed based on high
346 Cook's distance ($> \frac{4}{n}$). The color gradient and black dashed line correspond to data density and a
347 one-to-one relationship, respectively.

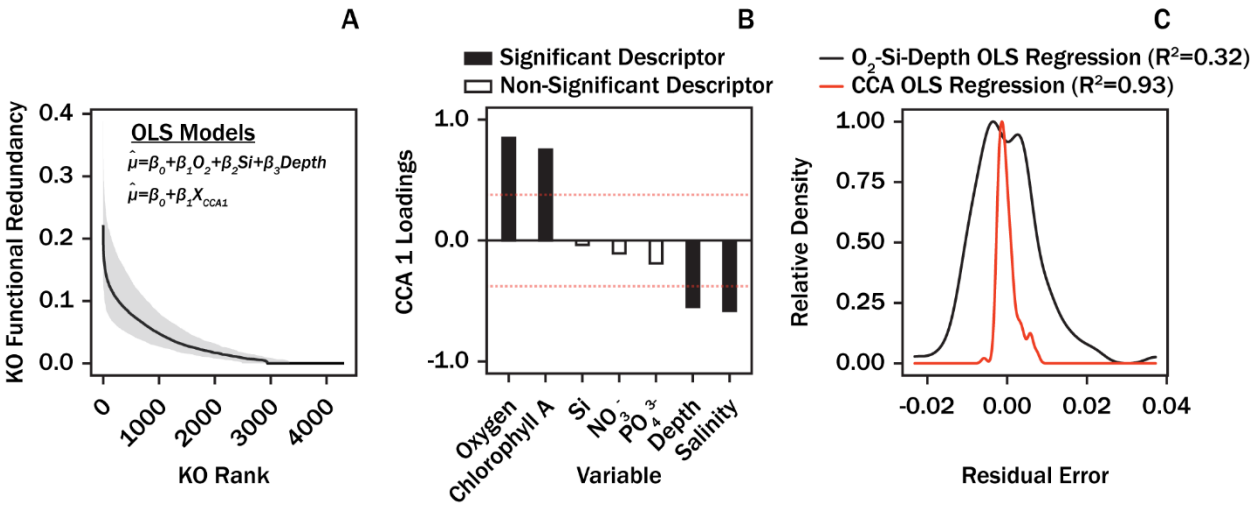
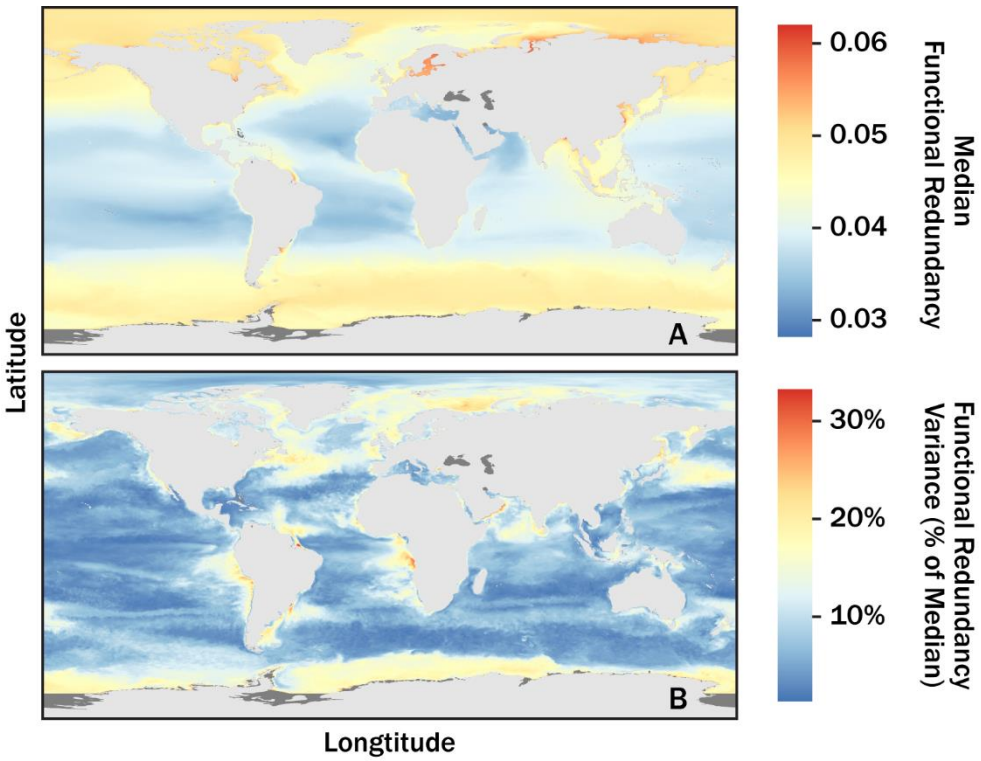


Fig. 2: Metatranscriptome rank-functional redundancy curves across all *TARA Oceans* sites (n=129) analyzed in this study (**A**). The black line corresponds to the median while the shaded area corresponds to the range spanning between the 5th and 95th percentiles. Proposed models were evaluated for accuracy in predicting mean (μ) and standard deviation (σ) of individual rank-functional redundancy curves. The O_2-Si-Depth model was selected as a best subset (minimum AIC) among the seven predictor variables. Canonical axis 1 loadings (scaling = 0) derived from the redundancy analysis (**B**). Dashed red lines in panel (**B)** correspond to the "equilibrium line of descriptors" ($\sqrt{d/p} = \sqrt{1/7}$), a threshold for defining significant variable contribution to factor loadings (*25*). A comparison in residual error of mean functional redundancy predicted by the O_2-Si-Depth OLS regression model and the canonical axis OLS regression model (**C**).



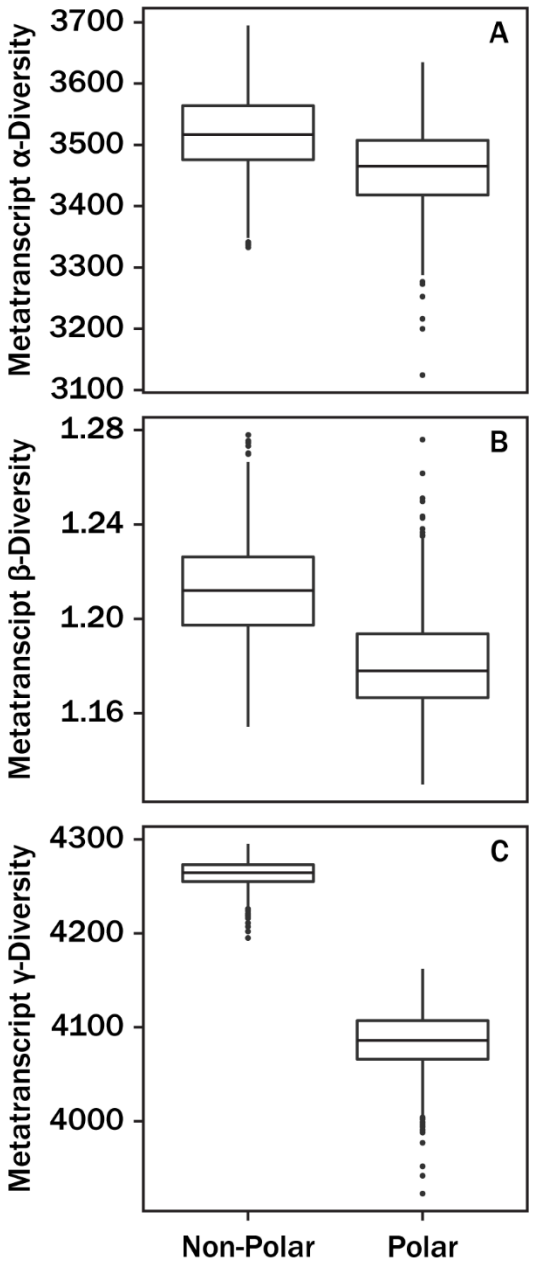
359

360 **Fig. 3:** Mean functional redundancy of all KOs at a 5m depth across Earth's oceans (0.25° x

361 0.25° resolution). Panels (**A**) and (**B**) correspond to the median and variance of predictions

362 spanning Jan 2013 to Dec 2018, respectively. Dark and light grey correspond to regions absent

363 of predictor data and land, respectively.



364 **Fig. 4.** Boxplots showing metatranscriptome α-diversity (**A**), β-diversity (**B**), and γ-diversity (**C**)
365 for polar and non-polar sites. Distributions are based 1,000 iterations of randomly sampling 10
366 metatranscriptomes, with replacement, from non-polar and polar latitudes.